RESEARCH ARTICLE

The impact is in the details: evaluating a standardized protocol and scale for determining non-native insect impact

Ashley N. Schulz[1], Angela M. Mech[2,15], Craig R. Allen[3], Matthew P. Ayres[4], Kamal J.K. Gandhi[5], Jessica Gurevitch[6], Nathan P. Havill[7], Daniel A. Herms[8], Ruth A. Hufbauer[9], Andrew M. Liebhold[10,11], Kenneth F. Raffa[12], Michael J. Raupp[13], Kathryn A. Thomas[14], Patrick C. Tobin[2], Travis D. Marsico[1]

1 *Arkansas State University, Department of Biological Sciences, PO Box 599, State University, AR 72467, USA* 2 *University of Washington, School of Environmental and Forest Sciences, 123 Anderson Hall, 3715 W Stevens Way NE, Seattle, WA 98195, USA* 3 *U.S. Geological Survey, Nebraska Cooperative Fish and Wildlife Unit, University of Nebraska-Lincoln, 423 Hardin Hall, 3310 Holdrege Street, Lincoln, NE 68583, USA* 4 *Dartmouth College, Department of Biological Sciences, 78 College Street, Hanover, NH 03755, USA* 5 *The University of Georgia, Daniel B. Warnell School of Forestry and Natural Resources, 180 E. Green St., Athens, GA 30602, USA* 6 *Stony Brook University, Department of Ecology and Evolution, Stony Brook, NY 11794, USA* 7 *USDA Forest Service Northern Research Station, 51 Mill Pond Rd., Hamden, CT 06514, USA* 8 *The Davey Tree Expert Company, 1500 N Mantua St., Kent, OH 44240, USA* 9 *Colorado State University, Department of Bioagricultural Science and Pest Management, 1177 Campus Delivery, Fort Collins, CO 80523, USA* 10 *USDA Forest Service Northern Research Station, 180 Canfield St., Morgantown, WV 26505, USA* 11 *Czech University of Life Sciences, Faculty of Forestry and Wood Sciences, Kamýcká 129 165 00 Praha 6–Suchdol, Prague, Czech Republic* 12 *University of Wisconsin, Department of Entomology, 1630 Linden Dr., Madison, WI 53706, USA* 13 *University of Maryland, Department of Entomology, 4112 Plant Sciences Building, College Park, MD 20742, USA* 14 *U.S. Geological Survey, Southwest Biological Science Center, 520 North Park St., Tucson, AZ 85719, USA* 15 *University of Maine, School of Biology and Ecology, 5751 Murray Hall, Orono, ME 04469, USA*

Corresponding author: Ashley N. Schulz (anschulz7@gmail.com)

Abstract

Assessing the ecological and economic impacts of non-native species is crucial to providing managers and policymakers with the information necessary to respond effectively. Most non-native species have minimal impacts on the environment in which they are introduced, but a small fraction are highly deleterious. The definition of 'damaging' or 'high-impact' varies based on the factors determined to be valuable by an individual or group, but interpretations of whether non-native species meet particular definitions can be influenced by the interpreter's bias or level of expertise, or lack of group consensus. Uncertainty or disagreement about an impact classification may delay or otherwise adversely affect policymaking on management strategies. One way to prevent these issues would be to have a detailed, nine-point impact scale that would leave little room for interpretation and then divide the scale into agreed upon categories, such as low, medium, and high impact. Following a previously conducted, exhaustive search regarding non-native, conifer-specialist insects, the authors independently read the same sources and scored the impact of 41 conifer-specialist insects to determine if any variation among assessors existed when using a detailed impact scale. Each of the authors, who were selected to participate in the working group associated with this study because of their diverse backgrounds, also provided their level of expertise and uncertainty for each insect evaluated. We observed 85% congruence in impact rating among assessors, with 27% of the insects having perfect inter-rater agreement. Variance in assessment peaked in insects with a moderate impact level, perhaps due to ambiguous information or prior assessor perceptions of these specific insect species. The authors also participated in a joint fact-finding discussion of two insects with the most divergent impact scores to isolate potential sources of variation in assessor impact scores. We identified four themes that could be experienced by impact assessors: ambiguous information, discounted details, observed versus potential impact, and prior knowledge. To improve consistency in impact decision-making, we encourage groups to establish a detailed scale that would allow all observed and published impacts to fall under a particular score, provide clear, reproducible guidelines and training, and use consensus-building techniques when necessary.

Keywords

environmental impact, expert opinion, impact assessment, joint fact-finding, non-native species management, policy-making, uncertainty

Introduction

Globally, anthropogenic, abiotic, and biotic threats increasingly affect the structure and function of forest ecosystems (Millar and Stephenson 2015). Of these threats, non-native species may cause considerable changes to the environments in which they are introduced, including ecological, economic, social, and cultural impacts (Jeschke et al. 2014). These impacts can be viewed as negative when there are undesirable effects or positive when they provide beneficial ecosystem services or economic value (Schlaepfer et al. 2011; Kumschick et al. 2012). Frequently, impacts must be assessed in the absence of sufficient published or otherwise available empirical data (Murray et al. 2009). One approach for estimating impact when empirical information is sparse (e.g., impacts on unclassified ecosystem services; Roy et al. 2018) is through surveys of expert opinion that consider the 'wisdom of the crowd' (e.g., observations, unpublished or preliminary datasets; Aspinall 2010; Gale et al. 2010; Thompson et al. 2013; Roy et al. 2014). However, it remains unclear how reliable expert opinion is.

In particular, consensus among experts may be difficult to achieve (Giannetti et al. 2009; Humair et al. 2014; González-Moreno et al. 2019). Further difficulty may occur when stakeholder groups and experts have different perspectives regarding the impact of non-native species. Disagreements and uncertainty among expert assessors, and between stakeholders and experts, may affect decision-making and resource allocation (Kumschick et al. 2012; Van Der Wal et al. 2015; Kumschick et al. 2015). For example, decision-makers may use information that is not necessarily based on taxon-specific scientific evidence, but rather broad ecological principles based on legal or regulatory considerations found in procedural manuals and technical guides developed by regulatory agencies (Fleischman and Briske 2016). This lack of taxon-specific, science-based evidence in the decision-making process may complicate the development and implementation of effective biosecurity policies, including surveillance and intervention strategies (Green et al. 2015).

Although disagreements may arise, impact assessments perform a crucial role in biosecurity programs for management of non-native species (Perrings et al. 2005; Hulme 2011). Many scales and assessment protocols have been developed to assess the impacts of non-native species on local or regional economies and societies. While new protocols, such as the INvasive Species Effects Assessment Tool (INSEAT; Martinez-Cillero et al. 2019), are being developed, some researchers are now evaluating the efficiency and efficacy of other long-standing impact assessment protocols to develop more robust, accurate, and consistent protocols. For example, González-Moreno et al. (2019) summarized and evaluated consistency in 11 commonly used protocols developed and applied in Europe, and found considerable inconsistency among assessors. Difficulties in creating and utilizing these standardized scoring systems and impact assessment protocols may include: 1) disagreement in how impact should be evaluated; 2) differences among the diverse array of introduced species and their typical and maximum impacts; 3) the extent to which species are broadly distributed versus limited to cultivated systems; and 4) differential impacts for unclassified ecosystem services and various socio-economic sectors (Humair et al. 2014; Roy et al. 2018). Consequently, experts often do not provide consistently defined impacts of studied organisms (Jeschke et al. 2014).

To help remedy inconsistency and disagreement among assessors, standard impact scoring systems (Kumschick et al. 2012; Blackburn et al. 2014; Roy et al. 2018) with seven to ten-points are suggested because they are more reliable and better measure an assessor's true evaluation (Preston and Colman 2000). Some impact scoring systems and assessment protocols have been developed in a way that can only be used by assessors with a high level of expertise as they require specialized knowledge about the species in question (González-Moreno et al. 2019). Other researchers argue that a diverse group of experts with broader knowledge should complete the assessments (e.g., Murray et al. 2009; Hemming et al. 2018a,b) to achieve accurate and consistent decisions. Additionally, structured protocols can help reduce biases and improve accuracy and transparency, and discussions can help resolve disagreements (Hemming et al. 2018a,b; González-Moreno et al. 2019; Osunkoya et al. 2019). This method of resolving conflicting assessments by allowing the assessors to openly discuss available data and the research used

to draw conclusions is known as joint fact-finding (Matsuura and Schenk 2016). Even with disagreements, the aggregated scores of a group tend to be closer to the true value than the score provided by any individual within the group (Roy et al. 2014).

Impact scores were recently used to categorize non-native forest insects that specialize on conifers (Mech et al. 2019a). During this project, a group of scientists (the "High-Impact Insect Invasion" working group; HIWG) collaborated to create a detailed nine-point scale of impact, but only one assessor was responsible for determining the impact score for the 58 non-native conifer-specialists currently in North America. These scores were eventually used as the basis for a statistical model that will be used to predict the impact of non-native conifer-specialists that have not yet become established in North America (Mech et al. 2019a). The purpose of our study was to evaluate whether the impact scale used in Mech et al. (2019a) is detailed enough for multiple people with different levels of expertise to reach the same impact score. We examined how level of expertise, uncertainty, and disagreement may affect impact assessment of non-native conifer-specialist insect species. Specifically, the objectives of the study were to: 1) evaluate the level of consensus among individual assessments of non-native insect impacts; 2) measure correlation among level of prior expertise, impact score, and assessor level of uncertainty; 3) assess the points of agreement and disagreement to determine which types of insects are the most difficult to assess with consensus; and 4) explore how experts can use joint fact-finding, a form of consensus-building, to identify sources of highly divergent impact scores and achieve consensus in decision-making using a case study of two insect species with highly divergent impact scores.

Materials and methods

Assessor group

In 2016, the HIWG, composed mainly of the co-authors of this paper, convened to examine the drivers of non-native insect invasions (Mech et al. 2019a) and develop a model to predict future high-impact, non-native, phytophagous insect species in natural ecosystems in North America. The group of scientists had different specialties (Suppl. material 1: Table S1) and diverse backgrounds (e.g., ethnic, cultural, age, stage of scientific career), with many having long-standing research experience in invasion ecology. Fifteen members of the 2016 HIWG participated in this project to determine whether the impact scores used in the analyses would be the same regardless of which working group member conducted the assessment.

Impact scoring system

The HIWG designed an original nine-point scale (Fig. 1) to classify the impacts of non-native insects already in North America (Mech et al. 2019a). We designed an original
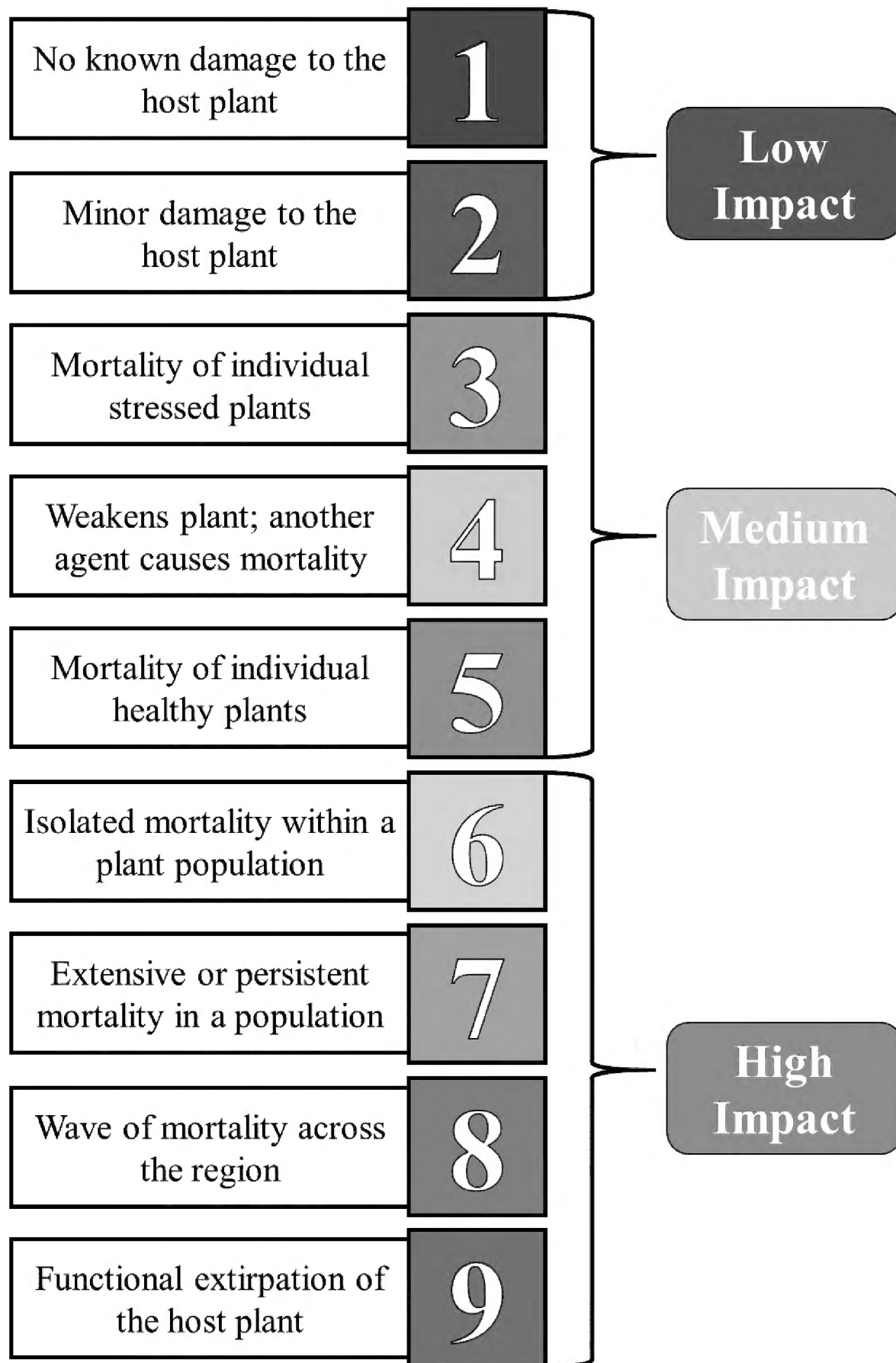
Figure 1. Impact scoring system (1–9) established by the HIWG for non-native, phytophagous insects, where 1 indicates no damage and 9 indicates severe damage (also found in Mech et al. 2019a).

scale because other impact scales were considered too general (e.g., EPPO-EIA, which addresses impacts of non-native plants and invertebrates overall), too specific (e.g., only addresses species within a particular feeding guild or region), or too complex (e.g., Kum-schick et al. 2015, the generic impact scoring system) for the primary purposes of the

project (i.e., Mech et al. 2019a). Our original impact scale ranged from 1–9, with one being the lowest and nine being the highest possible impact (Fig. 1). The HIWG determined that insects in levels 1–2 can be considered low impact species on a ternary impact scale (i.e., low, medium, or high), since they have no or minor (e.g., leaf or needle loss, foliage discoloration, twig dieback, cone drop) documented damage to their host plant. Insects in levels 3–5 can be considered medium impact species, since they cause mortality to individual host plants, and insects in levels 6–9 can be classified as high-impact because they cause mortality within a population of host plants (Fig. 1). The details in this scale were included with the goal that any description of impact in the literature would be able to fall under one of these scores (i.e., little need for interpretation).

Impact assessment

The HIWG initiated their research by conducting a pilot study on the 58 non-native, conifer-specialist insect species (i.e., restricted to feeding on one or more of the three conifer families in North America: Cupressaceae, Pinaceae, and Taxaceae) currently in North America (Mech et al. 2019a; Suppl. material 1: Table S2). For each non-native insect included in Mech et al. (2019a), one initial assessor conducted a comprehensive search of the peer-reviewed and gray literature (e.g., university and federal government websites, other credible online resources) to find any and all descriptions of impact. Gray literature was only referenced when publications were lacking, which typically occurred with insects that caused little to no damage. For each insect included in the study, the assessor identified the highest impact the insect had on trees native to North America. This information on the highest observed impact was used to determine impact score for each insect, and was used to create the models developed in Mech et al. (2019a).

For this study, we were interested in evaluating the impact scale used in Mech et al. (2019a), so we also focused on non-native, conifer-specialist insects in North America. For each conifer-specialist insect, assessors were provided with the list of references that described the host damage used to determine the impact scores used in Mech et al. (2019a). Of the 58 conifer-specialist insects that were originally identified in the pilot study, 17 insect species were excluded from our study because they received an impact score of one. This meant there was no documented damage and, therefore, no references were provided. The remaining 41 conifer-specialist insects (Suppl. material 1: Table S2; Fig. 2) were randomly assigned to three new assessors for impact scoring. In total, each insect was assessed by four assessors, including the original assessor who assessed the impacts for Mech et al. (2019a). The HIWG provided a diverse group to participate in the assessment (as suggested in Turbé et al. 2017 and Hemming et al. 2018a, b).

For each insect, the three new assessors were provided the same list of references as the initial assessor. The new assessors did not have access to the impact score assigned by the initial assessor to avoid bias. The references provided for each insect were mostly exhaustive, but for well-studied species (e.g., hemlock woolly adelgid [*Adelges tsugae* Annand]), references that were representative of the damage repeatedly found in published articles were selected in lieu of providing all impact literature. No publications
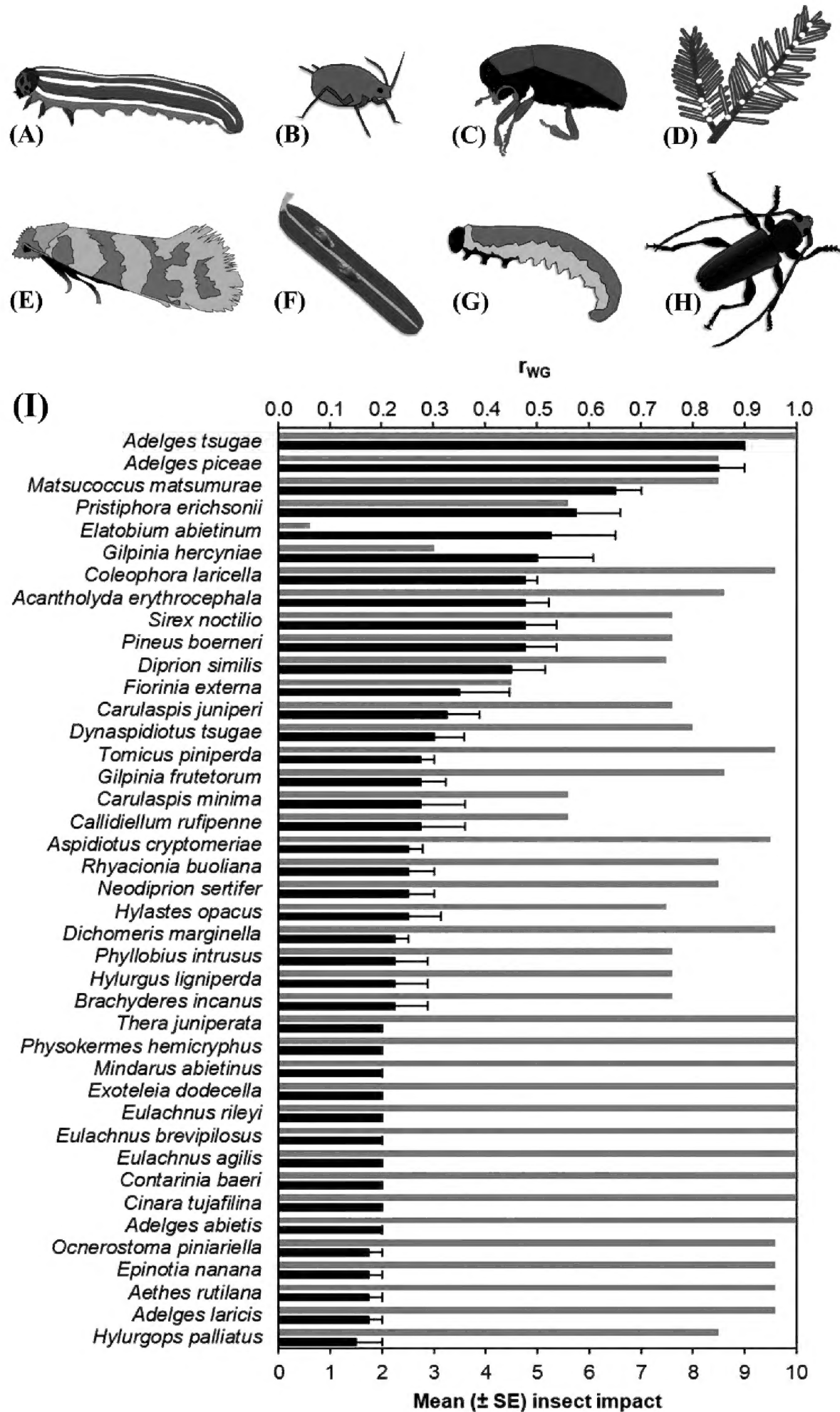
Figure 2. Examples of non-native, conifer-specialist insects, including (**A**) European spruce sawfly (*Gilpinia hercyniae*), (**B**) spruce needle aphid (*Elatobium abietinum*), (**C**) lesser spruce shoot beetle (*Hylurgops palliatus*), (**D**) hemlock woolly adelgid (*Adelges tsugae*), (**E**) pale juniper webworm (*Aethes rutilana*), (**F**) elongate hemlock scale (*Fiorinia externa*), (**G**) larch sawfly (*Pristiphora erichsonii*), and (**H**) Japanese cedar longhorned beetle (*Callidiellum rufipenne*) with the (**I**) mean (± SE) insect impact score (black bars) and within-group interrater agreement index (r_{WG}, gray bars) for all 41 conifer-specialist insects assessed in this study.

or websites, other than the ones provided, could be used by the assessors. Further, assessors were advised to not use their existing knowledge to evaluate impact and base their impact score solely on the information provided in the references.

Prior to completing the impact assessment exercise, assessors were provided with a sample score sheet that was developed by the first author. The score sheet included directions on how to assess impact and self-assign their level of expertise and uncertainty for each insect (Suppl. material 1: Table S3). Assessors were directed to select the highest applicable impact value based on their interpretation of the references. If a reference cited the impact of the insect on a conifer outside of North America, even if the conifer was native to North America, the assessors were instructed to disregard that information and only focus on the impacts that occurred in North America. For each insect, the assessors, including the initial assessor, self-reported their level of expertise on the insect they were assessing (scale of 1–5, from no to high expertise), as well as the level of uncertainty about their impact score decision (scale of 1–5, from low to high uncertainty) (Suppl. material 1: Table S3). During a conference call, assessors were trained to conduct an impact assessment using a sample insect not included in this study, and were given the opportunity to discuss any questions or concerns (approach also implemented by González-Moreno et al. 2019). Once all assessors were trained, score sheets with randomly assigned insects (from the list of 41 conifer-specialist insects; Suppl. material 1: Table S2) were sent to each assessor. Completed score sheets were assessed for completeness and then compiled into one spreadsheet with masked assessor identities.

Statistical analyses

Descriptive statistics were calculated for impact score and assessor levels of expertise and uncertainty for each insect, with all means reported ± 1 SE. A power function analysis was used to determine the required number of assessments per species. To evaluate the overall level of consensus among assessors, we calculated Krippendorff's alpha (Kα), a coefficient used to measure agreement among observers (Krippendorff 2017). To calculate Kα, we used the *kripp.alpha* function in the IRR (Interrater Reliability) package in R v.3.4.0 (R Core Team 2017; Gamer et al. 2012). Kα ranges from 0 to 1, with higher values indicating stronger agreement. In general, any values above 0.70 are thought to indicate high agreement (LeBreton and Senter 2008). To quantify agreement among the ordinal impact scores for each insect, we used the within-group inter-rater agreement index of r_{WG},

$$r_{WG} = 1 - \frac{s_X^2}{\sigma_E^2},$$

where s_x^2 is the observed variance among the impact scores from the four assessors, and σ_E^2 is the expected variance in the case of no consensus among assessors (LeBreton and Senter 2008).

When assessors are in perfect agreement, the index r_{WG} equals one, and any disagreement will cause the r_{WG} index to approach zero. Like Kα, r_{WG} = 0.70 is the traditionally accepted threshold that demarcates high versus low assessor agreement, whereby any values ≥ 0.70 indicate high agreement among assessors (LeBreton and Senter 2008). We used r_{WG} values to determine which insects were the most difficult to assess.

Spearman's rank correlation tests were conducted to measure the correlations between assessor levels of expertise and uncertainty. To measure whether expertise and uncertainty influence assignment of impact scores, we calculated the coefficients of variation for insect impact score, level of expertise, and level of uncertainty using the four assessor scores and ratings for each insect. We then conducted Spearman's rank correlation tests using the coefficients of variation for level of expertise and impact score and level of uncertainty and impact score, respectively.

Joint fact-finding meeting

Following the completion and compilation of all assessments, assessors met in person for a joint fact-finding session in August 2017 to identify potential sources of variation for insects with highly divergent impact scores. For our joint fact-finding discussion (Matsuura and Schenk 2016), we selected two conifer-specialist insects with the most divergent impact scores (i.e., lowest r_{WG} values): European spruce sawfly (*Gilpinia hercyniae* Hartig; Fig. 2A) and spruce needle aphid (*Elatobium abietinum* Walker; Fig. 2B). Since only four assessors evaluated these insects, references for the two species were provided to the group to read in preparation for the discussion. During this meeting, members reflected on the variance among impact scores for both insects and identified potential sources of uncertainty in the assessment of these insects.

Results

Mean impact scores ranged from 1.5 ± 0.5 for lesser spruce shoot beetle (*Hylurgops palliatus* Gyllenhal; Fig. 2C) to 9.0 ± 0.0 for hemlock woolly adelgid (Fig. 2D) (Table 1; Fig. 2). Although we removed 17 species that had an impact score of one (i.e., no documented damage) before the assessment, 12 of the remaining 41 insects that were evaluated had at least one assessor who scored the impact level as one. As a result, five insects (e.g., pale juniper webworm [*Aethes rutilana* Hübner; Fig. 2E]), had a mean impact score < 2. The coefficient of variation for impact score ranged from 0 to 67%, with 11 insects (27% of the insect species evaluated) having no variation in assessed impact scores (Fig. 2I; Fig. 3). The coefficient of variation peaked for insects with medium impact (levels 3–6), with less variation in extreme impact scores (i.e., high or low impact). We determined that, with four assessments per species, differences were readily evident among the 41 insects ($F_{40,123}$ = 11.49, P < 0.0001), and SE for species-specific estimates was approximately 0.53 on the nine-point scale of impact (Suppl. material 1: Fig. S1). The 95% CI with four assessors was ± 1.69 units on the nine-point scale.

Table 1. Summary of descriptive statistics (mean ± SE) for the self-assessed level of expertise (range of 1–5, in which 1 is no expertise and 5 is high expertise), impact level (range of 1–9, in which 1 is no documented damage and 9 is functional extinction of the host plant), and self-assessed level of uncertainty (scale of 1–5, where 1 is low uncertainty and 5 is high uncertainty) for each insect species assessed in this study.

| Conifer-specialist Insect Species	Mean ± SE Expertise	Mean ± SE Impact	Mean ± SE Uncertainty
Acantholyda erythrocephala	1.50 ± 0.50	4.75 ± 0.48	2.75 ± 0.25
Adelges abietis	2.75 ± 0.75	2.00 ± 0.00	1.50 ± 0.29
Adelges laricis	2.75 ± 0.85	1.75 ± 0.25	1.75 ± 0.25
Adelges piceae	3.75 ± 0.95	8.50 ± 0.50	2.00 ± 0.70
Adelges tsugae	4.75 ± 0.25	9.00 ± 0.00	1.50 ± 0.29
Aethes rutilana	2.00 ± 0.41	1.75 ± 0.25	1.75 ± 0.48
Aspidiotus cryptomeriae	2.75 ± 0.48	2.50 ± 0.29	2.00 ± 0.41
Brachyderes incanus	2.00 ± 0.00	2.25 ± 0.63	2.50 ± 0.65
Callidiellum rufipenne	1.50 ± 0.29	2.75 ± 0.85	2.25 ± 0.75
Carulaspis juniperi	2.50 ± 0.87	3.25 ± 0.63	2.50 ± 0.96
Carulaspis minima	1.75 ± 0.25	2.75 ± 0.85	3.00 ± 0.71
Cinara tujafilina	2.25 ± 0.25	2.00 ± 0.00	1.75 ± 0.25
Coleophora laricella	2.00 ± 0.41	4.75 ± 0.25	1.75 ± 0.48
Contarinia baeri	2.00 ± 0.41	2.00 ± 0.00	2.25 ± 0.25
Dichomeris marginella	2.00 ± 0.71	2.25 ± 0.25	2.75 ± 0.75
Diprion similis	2.25 ± 0.63	4.50 ± 0.65	2.50 ± 0.29
Dynaspidiotus tsugae	2.25 ± 0.25	3.00 ± 0.58	3.00 ± 0.71
Elatobium abietinum	2.00 ± 0.41	5.25 ± 1.25	1.75 ± 0.25
Epinotia nanana	1.25 ± 0.25	1.75 ± 0.25	2.25 ± 0.48
Eulachnus agilis	2.00 ± 0.41	2.00 ± 0.00	2.25 ± 0.75
Eulachnus brevipilosus	2.00 ± 0.41	2.00 ± 0.00	2.75 ± 0.85
Eulachnus rileyi	1.75 ± 0.25	2.00 ± 0.00	2.50 ± 0.65
Exoteleia dodecella	2.00 ± 0.41	2.00 ± 0.00	2.25 ± 0.48
Fiorinia externa	2.75 ± 1.03	3.50 ± 0.96	2.75 ± 0.75
Gilpinia frutetorum	2.50 ± 0.50	2.75 ± 0.48	2.25 ± 0.63
Gilpinia hercyniae	2.25 ± 0.48	5.00 ± 1.08	3.00 ± 0.58
Hylastes opacus	2.75 ± 0.75	2.50 ± 0.65	1.75 ± 0.48
Hylurgops palliatus	1.75 ± 0.25	1.50 ± 0.50	1.75 ± 0.48
Hylurgus ligniperda	1.50 ± 0.29	2.25 ± 0.63	2.50 ± 0.29
Matsucoccus matsumurae	2.50 ± 0.65	6.50 ± 0.50	2.75 ± 0.48
Mindarus abietinus	2.00 ± 0.41	2.00 ± 0.00	2.25 ± 0.63
Neodiprion sertifer	2.75 ± 0.85	2.50 ± 0.50	2.00 ± 0.41
Ocnerostoma piniariella	2.00 ± 0.41	1.75 ± 0.25	2.25 ± 0.48
Phyllobius intrusus	2.00 ± 0.41	2.25 ± 0.63	2.25 ± 0.48
Physokermes hemicryphus	2.00 ± 0.41	2.00 ± 0.00	2.25 ± 0.25
Pineus boerneri	3.25 ± 0.85	4.75 ± 0.63	2.50 ± 0.65
Pristiphora erichsonii	2.50 ± 0.87	5.75 ± 0.85	2.00 ± 0.41
Rhyacionia buoliana	3.25 ± 0.48	2.50 ± 0.50	1.75 ± 0.25
Sirex noctilio	2.50 ± 0.96	4.75 ± 0.63	2.00 ± 0.41
Thera juniperata	2.00 ± 0.41	2.00 ± 0.00	2.25 ± 0.48
Tomicus piniperda	3.00 ± 0.71	2.75 ± 0.25	1.50 ± 0.29

The r_{WG} index to assess within-group variation for each species varied from 0.06–1.00, with 85% (35 of 41) of the insects having a $r_{WG} \geq 0.70$ and 27% (11 out of 41) having a $r_{WG} = 1.00$ (Fig. 2I). The 11 species with perfect agreement (those with no variation) had a mean impact of 2, except hemlock woolly adelgid, which had a mean
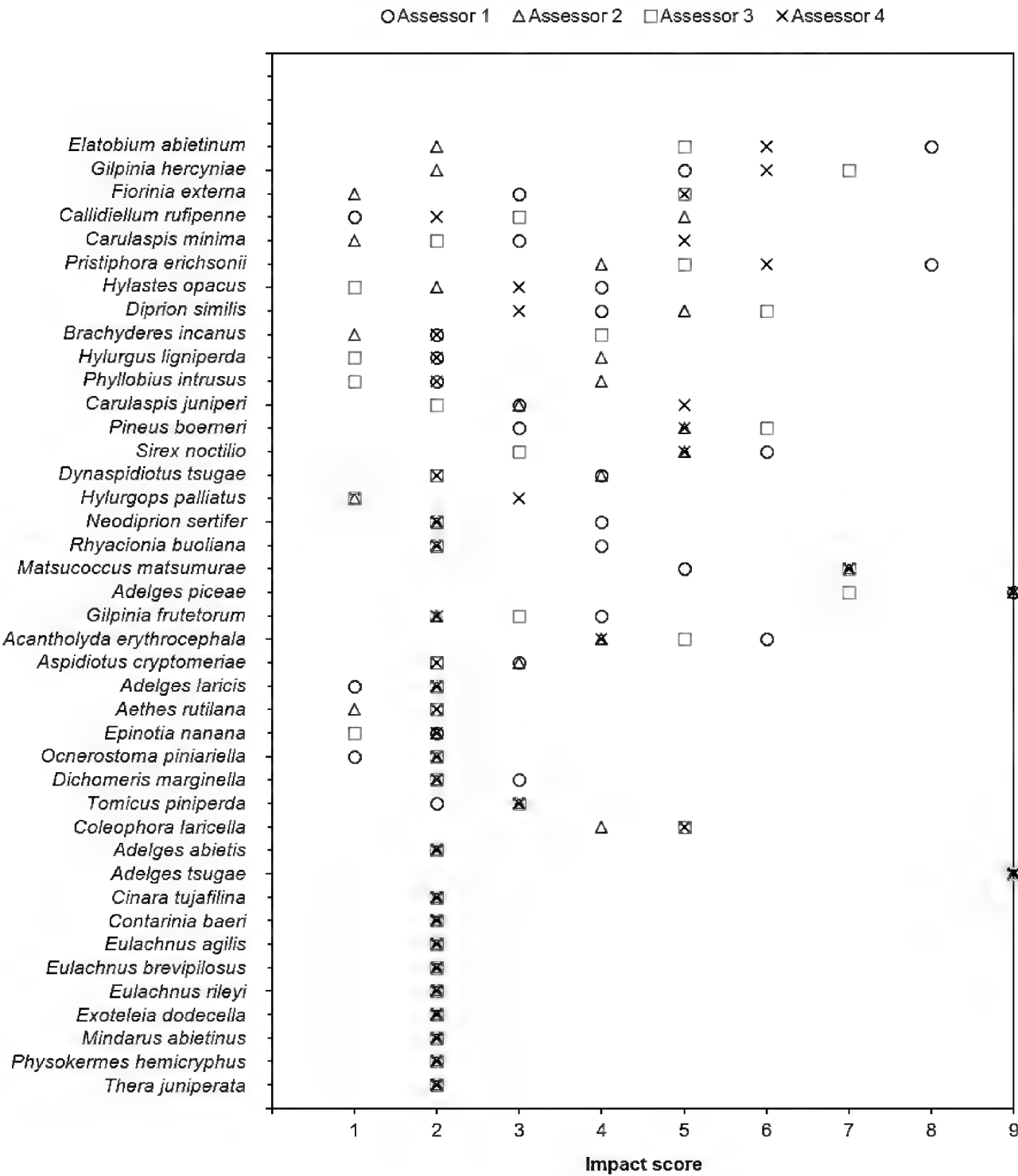
Figure 3. Insect impact scores assigned by each of the four assessors for each insect. Insects with the most disagreement are at the top of the figure, whereas insects with the most consensus are at the bottom of the figure.

impact of 9 (Fig. 3). As with the coefficient of variation, insects with a medium impact tended to exhibit the most divergence in assessed values among experts ($r_{WG} < 0.70$; Fig. 4). The mean impact score of the six species (15% of those in the sample) generating the most disagreement ($r_{WG} < 0.70$) ranged from 2.75–5.75 (Figs 3, 4). These include elongate hemlock scale (*Fiorinia externa* Ferris; Fig. 2F), European spruce sawfly, larch sawfly (*Pristiphora erichsonii* Hartig; Fig. 2G), Japanese cedar longhorned beetle (*Callidiellum rufipenne* Motschulsky; Fig. 2H), minute cypress scale (*Carulaspis minima* Borchsenius), and spruce needle aphid. For all 41 insect impact assessments, Kα was 0.55.

The mean self-assessed level of expertise ranged from 1.25 ± 0.25 (novice; no expertise) for European spruce needle miner (*Epinotia nanana* Treitschke) to 4.75 ± 0.25 (expert; high expertise) for hemlock woolly adelgid (Table 1; Suppl. material 1: Fig.
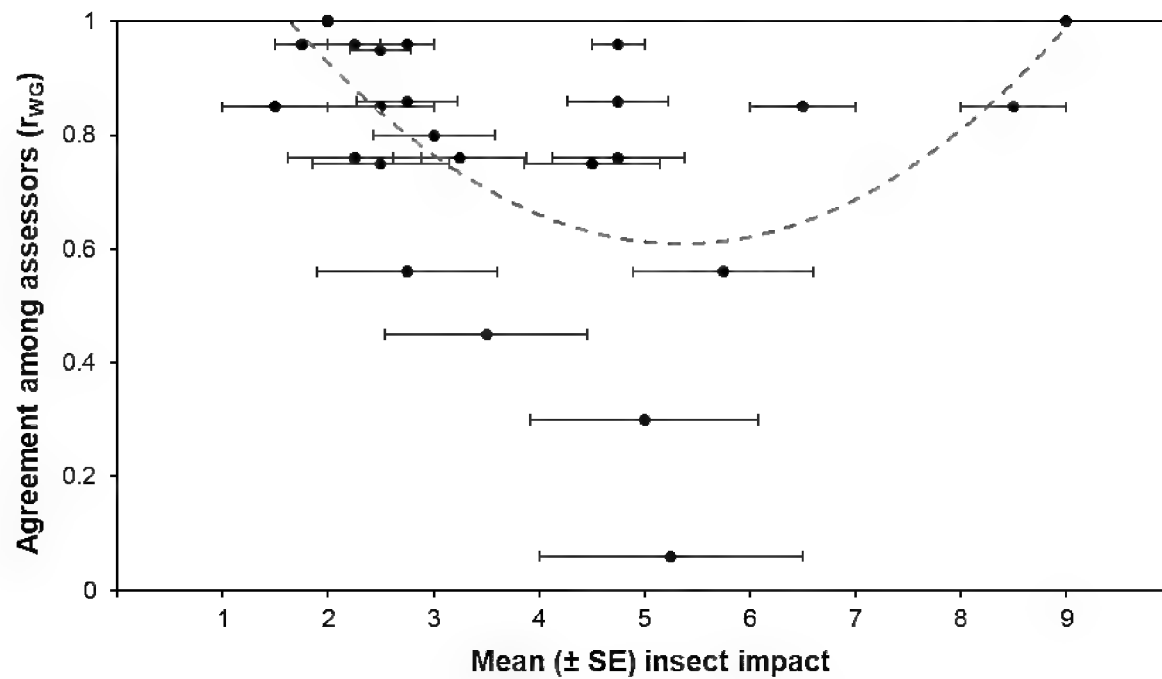
Figure 4. Within-group inter-rater agreement (r_{WG}) values (0–1, with 0 indicating no agreement and 1 indicating perfect agreement) for each mean (± SE) insect impact (1–9, with 1 indicating low impact and 9 indicating high impact insect species) with a trendline shown in red.
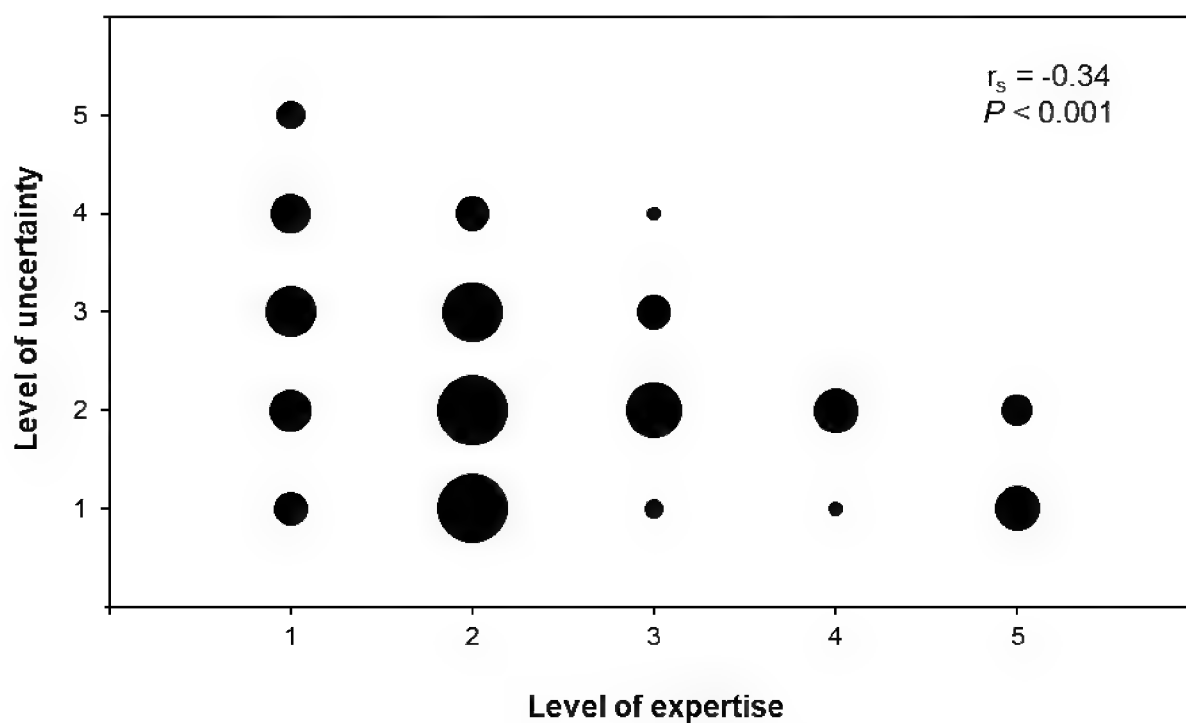
Figure 5. Significant, positive correlation between level of expertise (scale of 1–5, from no to high expertise) and level of uncertainty (scale of 1–5, from low to high uncertainty) with bubbles that are proportional to the number of overlapping data points.

S2). The overall mean level of expertise for all 41 insects that were assessed was 2.3 ± 0.6 (advanced beginner; low expertise). The mean self-assessed level of uncertainty ranged from 1.5 ± 0.3 (no uncertainty) for eastern spruce gall adelgid, European pine shoot borer (*Tomicus piniperda* L.), and hemlock woolly adelgid to 3.0 ± 0.7 (moderate uncertainty) for minute cypress scale and shortneedle conifer scale (*Dynaspidiotus tsugae* Marlatt) (Table 1). The overall mean level of uncertainty for all 41 insect assessments was 2.2 ± 0.5 (low uncertainty). The levels of expertise and uncertainty were

Table 2. Common themes that emerged from the joint fact-finding discussion on variation in non-native, conifer-specialist insect impact scores and reflection on problems that the assessors encountered when making their assessments.

Theme	Description
Ambiguous information	Information in the literature was vague, lacking, incorrect, or unconvincing. Often, very little information was provided on the impacts of generally low impact species. Misinterpretation of the ambiguous information provided in the references may have resulted in an under- or over-estimated impact score.
Discounted details	The assessor unintentionally overlooked details because s/he did not thoroughly read the provided literature. Alternatively, the assessor may have intentionally disregarded details.
Observed vs. potential impact	Some references provided understated or overexaggerated impacts not supported by empirical data or observations. The assessor did not find it acceptable to assign a lower or higher impact when the species had rarely achieved that potential.
Prior knowledge	A more specialized assessor had previous knowledge about the insect. Consequently, s/he had more insight than what was provided in the references and/or disagreed with the content in the references based on personal experiences with the insect.

negatively correlated (r_s = -0.34, P < 0.001, Fig. 5), whereas the correlations between the coefficients of variation for level of expertise and impact score (r_s = -0.05, P = 0.77) and level of uncertainty and impact score (r_s = 0.11, P = 0.49) were not significant.

The joint fact-finding discussion on European spruce sawfly and spruce needle aphid allowed the working group to constructively reflect on the variation in insect impact scores and identify potential sources of uncertainty. The joint fact-finding meeting also provided a forum to discuss problems that assessors encountered when assigning impact scores for other insects included in this study. Four common themes emerged from the discussion: ambiguous information, discounted details, observed vs. potential impact, and prior knowledge (Table 2). The group discussed and resolved divergent impact scores, concluding the meeting with participant agreement that both the European spruce sawfly and spruce needle aphid should be assigned level 6 on the nine-point impact scale.

Discussion

Impact assessment protocols for non-native insect species

For this study, we evaluated the efficacy of a detailed nine-point impact scale (Fig. 1) that was developed to assess impacts of non-native insects in forests. Our decision to only have four assessors score each insect rather than every assessor score each insect was supported by the results of our power function analysis (Suppl. material 1: Fig. S1). Employing four assessments per insect species allowed us to evaluate many species while still having reasonable precision in the species-specific estimates.

We found 11 of the 41 non-native, conifer-specialist insects assessed had perfect agreement among assessors, 24 had a high level of agreement, and only six elicited a low level of agreement. Although the Krippendorff's alpha indicated a moderate level of consensus, the fact that most insects had a high or perfect level of agreement indicated

a generally high consensus among assessors. All insects with low agreement among assessors were scored within or on the margin of the medium impact range, whereas the insects with perfect or high agreement among assessors fell near the extremes of their respective impact range. This pattern indicates that divergence in agreement peaked in insects with a medium impact score, perhaps highlighting the challenges associated with determining impact for species that are neither truly benign (low-impact) nor undeniably catastrophic (high-impact). Our use of standardized information may have contributed to this pattern, as this limited the information assessors used to make their assessment. The initial assessor endeavored to select the most comprehensive and accurate references available, but published information can be vague, inaccurate, or misinterpreted. Although we advised assessors to not use their prior knowledge, some assessors had specialized expertise to use when the literature was deficient, while others disagreed with what was written. The joint fact-finding discussions improved understanding and ultimately led to consensus about these medium-impact species. Following the discussions and reassessment, there was no variability in which impact level (low, medium, or high) all 41 insects should be.

This pattern of highly divergent impact scores may also result from intraspecific variation in impact. For this assessment, we considered a taxonomic definition of impact (Colautti and MacIsaac 2004; i.e., a species manifests the same level of impact throughout its invaded region). However, a medium score could reflect regional variation in impact. For example, one population may have natural enemies that limit impact, whereas another population does not. Regional variation in impact score may also reflect differences in stakeholder perceptions, as individuals living in urban areas may perceive impact to be higher, whereas people in rural areas may perceive impact to be lower (Kumschick et al. 2012; Jeschke et al. 2014). Although we advised assessors to select the highest impact score supported by the information in the provided literature, some assessors may have overlooked details about intraspecific variation in impact or assigned an average score that considered the impacts in all of the regions.

Higher variation among medium impact species highlights the importance of having a robust impact scoring system. Although a few impact assessment scoring systems have multiple levels with detailed descriptions from which to choose (e.g., Ricciardi and Cohen 2007; D'hondt et al. 2015; Nentwig et al. 2016), most impact assessment protocols employ an impact scale with three to five levels (e.g., Kenis et al. 2012; Martinez-Cillero et al. 2019). Overall, the generally high level of consensus in our assessment may be attributed in part to our clearly defined impact scoring system.

Assessor expertise and uncertainty

In this study, the overall self-assessed expertise level was low, with most insects eliciting an expertise level below three (moderate expertise). The only species that elicited a moderate-high to high self-assessed expertise (> level 3 on the expertise scale) were high impact species: balsam woolly adelgid (*Adelges piceae* Ratzeburg), hemlock woolly adelgid, and pine woolly aphid (*Pineus boerneri* Annand). In a pool of assessors, one would

expect to have more assessors with expertise on high- than low-impact insect species because high-impact species generate more research funding and publicity in the academic community (e.g., more peer-reviewed publications) and the general public (e.g., more outreach and awareness efforts) than low-impact species. All three species are high-profile insects with widespread documentation, research, and public reporting, such that even non-specialist scientists may be acquainted enough with these species to rate their expertise level as high. High self-assessed levels of expertise might also be elicited from other high-impact species not included in this study.

Uncertainty is often of concern when assessing impact. It is important for assessors to consider the available information and determine the potential impact that the non-native species has or will have with accuracy and consistency to efficiently allocate resources to management and biosecurity strategies (Andersen et al. 2004). In our study, the level of self-assessed uncertainty was low, with all insects eliciting a self-assessed uncertainty level of ≤ 3. In other words, most assessors were confident in their decisions. This confidence could be attributed, in part, to our simple, yet clearly defined impact scoring system, which reduced the need for complex interpretation and guessing. Achieving consistent decision with certainty is often difficult. In situations where assessors have uncertainty not eliminated with appropriate elicitation and consensus-building techniques (e.g., lack of data or uneven evidence base), it has been suggested that assessors should quantify and communicate their true level of uncertainty to decision-makers for use in the decision-making process (Aspinall et al. 2010; Turbé et al. 2017; Vanderhoeven et al. 2017). Assessors can abide by the precautionary principle (Kriebel et al. 2001) and consider the species a higher risk until more information can be collected to indicate otherwise (Vanderhoeven et al. 2017).

Most studies that address expertise and expert opinion also address uncertainty (e.g., Murray et al. 2009; Vanderhoeven et al. 2017; Roy et al. 2018; González-Moreno et al. 2019) because the two variables can be closely associated. We observed a negative correlation between these two variables (Fig. 5), indicating assessors with high levels of expertise were more certain than assessors with lower levels of expertise. This pattern may be expected if experts generally have more prior knowledge, making them more certain. However, our assessors self-assigned fairly low levels of expertise and uncertainty, which is seemingly inconsistent with the negative correlation we observed. Many assessors rated themselves as "low expertise-no uncertainty" and "low expertise-low uncertainty" rather than "no expertise" (Fig. 5), which may have contributed to the negative correlation.

We observed no associations between the coefficient of variation for impact score and the coefficients of variation for the levels of expertise and uncertainty, as both correlations were non-significant. This suggests that expertise and uncertainty may not influence the interpretation of non-native insect impact. In other words, assessors interpreted the same information and arrived at similar conclusions regardless of specific expertise. This is a good indication that the goal of the HIWG for designing the detailed impact scale was met—the same conclusions would most likely be met regardless of which group member did the assessing. It is worth noting that although assessors varied in their self-reported expertise, all are trained ecologists with experience interpreting ecological literature and may be considered "experts" as defined by Krueger et al. (2012).

Collaborative discussion promotes assessor consensus

Consensus-building and other participatory techniques are increasingly cited in the environmental impact assessment literature (e.g., Hemming et al. 2018b; González-Moreno et al. 2019; Osunkoya et al. 2019). Social scientists have long used approaches such as the Delphi technique, a process that uses iterative structured questionnaires and group communication to evaluate expert knowledge (e.g., Mukherjee et al. 2015), and general discussion (e.g., Hemming et al. 2018b; González-Moreno et al. 2019; Osunkoya et al. 2019) such as joint fact-finding (e.g., Matsuura and Schenk 2016). However, these techniques are still new to studies of biological invasions. Through our consensus-building discussion, we were able to identify four common themes regarding problems encountered by assessors when making their assessments (Table 2).

The first theme, *ambiguous information*, was a common problem encountered by the initial assessors as they sorted through the provided literature, much of which was vague or lacking. This problem was especially acute for species categorized as low impact, some of which were scored as level one, indicating that the new assessor read no information regarding impact, whereas the initial assessor documented at least minor damage. We determined that many of these errors were due to ambiguous language in the references (e.g., Jeschke et al. 2014) that may have led to misinterpretation of the information. Consensus-building discussions among expert assessors may help alleviate this problem.

The second theme that emerged regarded *discounted details*. Some of the sources referenced were lengthy and detailed, while others were more anecdotal and lacked sufficient detail for rigorous evaluation. An assessor that does not carefully read a reference in its entirety may overlook important details about impacts or the assessor may disregard some statements altogether. For example, an assessor may discount a specific older source because subsequent controlled experiments failed to replicate it. This source of variation may be alleviated if an assessor expresses concerns to the other expert assessors during discussion.

The third theme that emerged focused on *observed versus potential impacts*. Some references discussed potential impacts not yet supported by empirical data or observations and the assessor did not find it appropriate to assign a score based solely on this interpretation of potential. Our assessments were based on documented impacts rather than potential for future impacts (e.g., under predicted global climate change scenarios or once new hosts were accessed). Other impact assessment protocols, such as Sandvik et al. (2019), have established criteria for quantifying invasion potential of non-native species in all taxonomic groups. As with previous themes, this issue can be addressed through rating scale clarification and assessor consensus.

The final theme focused on variation from *prior knowledge*. In some cases, an assessor had more insight than provided in the references, but their perception differed little from the reference. In other scenarios, the assessor had experimental results or insight that did not support or failed to replicate the reference information, so they chose to base their score accordingly. Such decisions can contribute variation, whether or not

the assessor incorrectly rejects correct information. This scenario highlights the value of strict, standardized guidelines, and consensus-building techniques (Hemming et al. 2018b; González-Moreno et al. 2019; Osunkoya et al. 2019) that generate alternative perspectives guiding the group to a more uniform consensus.

Additional consensus was achieved through our joint fact-finding activity. The open dialogue among assessors facilitated achievement of consensus because assessors were able to critically evaluate ambiguous statements and, since some members of the group had prior knowledge that they used to inform their decisions, provide background knowledge based on experience not documented in the literature. Turbé et al. (2017) found that applying a similar joint fact-finding approach, along with clear guidelines and closed-ended questions, considerably improved outcomes. Other studies have also successfully used discussion groups to address uncertainty and disagreement and to make final decisions on environmental impacts (e.g., Hemming et al. 2018b; González-Moreno et al. 2019).

Conclusions

As written, the protocol and detailed, nine-point impact scale provided by the HIWG has the potential to result in a lack of consensus, particularly with medium-impact insect species. However, we found that adding joint fact-finding can alleviate any potential discrepancies in impact scoring. We demonstrate that consensus among diverse expert assessors can be achieved for invasive species decision-making and management. When empirical data are lacking for specific species, decision-makers may use broad ecological principles (Fleischman and Briske 2016) for management decisions, which is not ideal. To aid in the decision-making process, experts can first work independently to use rapid risk assessment techniques (e.g., Alves da Rosa et al. 2017) to characterize the impacts of the target species, after which consensus-building techniques can be used to reduce uncertainty and variation in impact scores (Turbé et al. 2017; Vanderhoeven et al. 2017; Hemming et al. 2018a,b; González-Moreno et al. 2019; Osunkoya et al. 2019). Reliable assessments based on vetted scientific evidence bolstered by diverse expert opinion and transparency about uncertainty (Vanderhoeven et al. 2017) will benefit decision-makers and managers tasked with allocating finite resources to manage the many threats confronting global ecosystems.

Data accessibility

All of the references used for this impact assessment are archived in the U.S. Geological Survey ScienceBase Catalog (Mech et al. 2019b). Suppl. material 1: Table S4 includes the level of expertise, impact score, and level of uncertainty assigned by the four assessors for each of the 41 conifer-specialist insects included in this study.

Acknowledgements

References

Alves da Rosa C, Zenni R, Ziller SR, de Almeida Curi N, Passamani M (2018) Assessing the risk of invasion of species in the pet trade in Brazil. Perspectives in Ecology and Conservation 16: 38–42. https://doi.org/10.1016/j.pecon.2017.09.005

Andersen MC, Adams H, Hope B, Powell M (2004) Risk assessment for invasive species. Risk Analysis 24: 787–793. https://doi.org/10.1111/j.0272-4332.2004.00478.x

Aspinall W (2010) A route to more tractable expert advice. Nature 463: 294–295. https://doi.org/10.1038/463294a

Blackburn TM, Essl F, Evans T, Hulme PE, Jeschke JM, Kühn I, Kumschick S, Marková Z, Mrugała A, Nentwig W, Pergl J (2014) A unified classification of alien species based on the magnitude of their environmental impacts. PLoS Biology 12: e1001850. https://doi.org/10.1371/journal.pbio.1001850

Colautti RI, MacIsaac HJ (2004) A neutral terminology to define 'invasive' species. Diversity and Distributions 10: 135–141. https://doi.org/10.1111/j.1366-9516.2004.00061.x

D'hondt B, Vanderhoeven S, Roelandt S, Mayer F, Versteirt V, Adriaens T, Ducheyne E, Martin GS, Grégoire J-C, Stiers I, Quoilin S, Cigar J, Heughebaert A, Branquart E (2015) Harmonia+ and Pandora+: risk screening tools for potentially invasive plants, animals and their pathogens. Biological Invasions 17: 1869–1883. https://doi.org/10.1007/s10530-015-0843-1

Fleischman F, Briske DD (2016) Professional ecological knowledge: an unrecognized knowledge domain within natural resource management. Ecology and Society 21: 32–40. https://doi.org/10.5751/ES-08274-210132

Gale P, Brouwer A, Ramnial V, Kelly L, Kosmider R, Fooks AR, Snary EL (2010) Assessing the impact of climate change on vector-borne viruses in the EU through the elicitation of expert opinion. Epidemiology and Infection 138: 214–225. https://doi.org/10.1017/S0950268809990367

Gamer M, Lemon J, Fellows I, Singh P (2012) Various coefficients of interrater reliability and agreement. R package *irr*, (v 0.84). https://cran.r-project.org/web/packages/irr/irr.pdf

Giannetti BF, Bonilla SH, Silva CC, Almeida CMVB (2009) The reliability of experts' opinions in constructing a composite environmental index: The case of ESI 2005. Journal of Environmental Management 90: 2448–2459. https://doi.org/10.1016/j.jenvman.2008.12.018

González-Moreno P, Lazzaro L, Vilà M, Preda C, Adriaens T, Bacher S, Brundu G, Copp GH, Essl F, García-Berthou E, Katsanevakis S, Moen TL, Lucy FE, Nentwig W, Roy HE, Srėbalienė G, Talgø V, Vanderhoeven S, Andjelković A, Arbačiauskas K, Auger-Rozenberg M-A, Bae M-J, Bariche M, Boets P, Boieiro M, Borges PA, Canning- Clode J, Cardigos F, Chartosia N, Cottier-Cook EJ, Crocetta F, D'hondt B, Foggi B, Follak S, Gallardo B, Gammelmo Ø, Giakoumi S, Giuliani C, Fried G, Jelaska LS, Jeschke JM, Jover M, Juárez-Escario A, Kalogirou S, Kočić A, Kytinou E, Laverty C, Lozano V, Maceda-Veiga A, Marchante E, Marchante H, Martinou AF, Meyer S, Michin D, Montero-Castaño A, Morais MC, Morales-Rodriguez C, Muhthassim N, Nagy ZA, Ogris N, Onen H, Pergl J, Puntila R, Rabitsch W, Ramburn TT, Rego C, Reichenbach F, Romeralo C, Saul W-C, Schrader G, Sheehan R, Simonović P, Skolka M, Soares AO, Sundheim L, Tarkan AS, Tomov R, Tricarico E, Tsiamis K, Uludağ A, van Valkenburg J, Verreycken H, Vettraino AM, Vilar L, Wiig Ø, Witzell J, Zanetta A, Kenis M (2019) Consistency of impact assessment protocols for non-native species. NeoBiota 44: 1–25. https://doi.org/10.3897/neobiota.44.31650

Green OO, Garmestani AS, Allen CR, Gunderson LH, Ruhl JB, Arnold CA, Graham NA, Cosens B, Angeler DG, Chaffin BC, Holling CS (2015) Barriers and bridges to the integration of social-ecological resilience and law. Frontiers in Ecology and the Environment 13: 332–337. https://doi.org/10.1890/140294

Hemming V, Burgman MA, Hanea AM, McBride MF, Wintle BC (2018a) A practical guide to structured expert elicitation using the IDEA protocol. Methods in Ecology and Evolution 9: 169–180. https://doi.org/10.1111/2041-210X.12857

Hemming V, Walshe TV, Hanea AM, Fidler F, Burgman MA (2018b) Eliciting improved quantitative judgements using the IDEA protocol: A case study in natural resource management. PloS One 13: e0198468. https://doi.org/10.1371/journal.pone.0198468

Hulme PE (2011) Biosecurity: the changing face of invasion biology. In: Richardson DM (Ed.) Fifty Years of Invasion Ecology: The Legacy of Charles Elton. Blackwell Publishing Ltd, West Sussex, 301–314. https://doi.org/10.1002/9781444329988

Humair F, Edwards PJ, Siegrist M, Kueffer C (2014) Understanding misunderstandings in invasion science: why experts don't agree on common concepts and risk assessments. NeoBiota 20: 1–30. https://doi.org/10.3897/neobiota.20.6043

Jeschke JM, Bacher S, Blackburn TM, Dick JTA, Essl F, Evans T, Gaertner M, Hulme PE, Kühn I, Mrugała A, Pergl J, Pyšek P, Rabitsch W, Ricciardi A, Richardson DM, Sendek A, Vilà M, Winter M, Kumschick S (2014) Defining the impact of non-native species. Conservation Biology 28: 1188–1194. https://doi.org/10.1111/cobi.12299

Kenis M, Bacher S, Baker RHA, Branquart E, Brunel S, Holt J, Hulme PE, MacLeod A, Pergl J, Petter F, Pyšek P, Schrader G, Sissons A, Starfinger U, Schaffner U (2012) New protocols to assess the environmental impact of pests in the EPPO decision-support scheme for pest risk analysis. EPPO Bulletin 42: 21–27. https://doi.org/10.1111/j.1365-2338.2012.02527.x

Kriebel D, Tickner J, Epstein P, Lemons J, Levins R, Loechler EL, Quinn M, Rudel R, Schettler T, Stoto M (2001) The precautionary principle in environmental science. Environmental Health Perspectives 109: 871–876. https://doi.org/10.1289/ehp.01109871

Krippendorff K (2017) Intercoder Reliability Techniques: Krippendorff's Alpha. In: Allen M (Ed.) The SAGE Encyclopedia of Communication Research Methods. SAGE Publishing Inc, Thousand Oaks, 743–751. https://doi.org/10.4135/9781483381411.n259

Krueger T, Page T, Hubacek K, Smith L, Hiscock K (2012) The role of expert opinion in environmental modelling. Environmental Modelling and Software 36: 4–18. https://doi.org/10.1016/j.envsoft.2012.01.011

Kumschick S, Bacher S, Dawson W, Heikkilä J, Sendek A, Pluess T, Robinson T, Kühn I (2012) A conceptual framework for prioritization of invasive alien species for management according to their impact. NeoBiota 15: 69–100. https://doi.org/10.3897/neobiota.15.3323

Kumschick S, Bacher S, Evans T, Markova Z, Pergl J, Pyšek P, Vaes-Petignat S, Veer G, Vilà M, Nentwig W (2015) Comparing impacts of alien plants and animals in Europe using a standard scoring system. Journal of Applied Ecology 52: 552–561. https://doi.org/10.1111/1365-2664.12427

LeBreton JM, Senter JL (2008) Answers to 20 questions about interrater reliability and interrater agreement. Organizational Research Methods 11: 815–852. https://doi.org/10.1177/1094428106296642

Martinez-Cillero R, Willcock S, Perez-Diaz A, Joslin E, Vergeer P, Peh KS (2019) A practical tool for assessing ecosystem services enhancement and degradation associated with invasive alien species. Ecology and Evolution 9: 3918–3936. https://doi.org/10.1002/ece3.5020

Matsuura M, Schenk T (2016) Joint Fact-Finding in Urban Planning and Environmental Disputes. Routledge, New York, 1–234. https://doi.org/10.4324/9781315651842

Mech AM, Thomas KA, Marsico TD, Herms DA, Allen CR, Ayres MP, Gandhi KJK, Gurevitch J, Havill NP, Hufbauer RA, Liebhold AM, Raffa KF, Schulz AN, Uden DR, Tobin PC. Evolutionary history predicts the impact of non-native herbivorous insects on North American conifers. Ecology and Evolution 9(21): 12216–12230. https://doi.org/10.1002/ece3.5709

Mech AM, Thomas KA, Havill NP, Schulz AN, Tobin PC (2019b) Traits and Factors Catalog (TRA-FAC): Conifer specialists of North America. U.S. Geological Survey ScienceBase (data release).

Millar CI, Stephenson NL (2015) Temperate forest health in an era of emerging megadisturbance. Science 349: 823–826. https://doi.org/10.1126/science.aaa9933

Mukherjee N, Huge J, Sutherland WJ, McNeill J, Van Opstal M, Dahdouh-Guebas F, Koedam N (2015) The Delphi technique in ecology and biological conservation: applications and guidelines. Methods in Ecology and Evolution 6: 1097–1109. https://doi.org/10.1111/2041-210X.12387

Murray JV, Goldizen AW, O'Leary RA, McAlpine CA, Possingham HP, Choy SL (2009) How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies *Petrogale penicillata*. Journal of Applied Ecology 46: 842–851. https://doi.org/10.1111/j.1365-2664.2009.01671.x

Nentwig W, Bacher S, Pyšek P, Vilà M, Kumschick S (2016) The generic impact scoring system (GISS): a standardized tool to quantify the impacts of alien species. Environmental Monitoring and Assessment 188: 315. https://doi.org/10.1007/s10661-016-5321-4

Osunkoya OO, Froese JG, Nicol S (2019) Management feasibility of established invasive plant species in Queensland, Australia: A stakeholders' perspective. Journal of Environmental Management 246: 484–495. https://doi.org/10.1016/j.jenvman.2019.05.052

Perrings C, Dehnen-Schmutz K, Touza J, Williamson M (2005) How to manage biological invasions under globalization. Trends in Ecology and Evolution 20: 212–215. https://doi.org/10.1016/j.tree.2005.02.011

Preston CC, Colman AM (2000) Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychologica 104: 1–15. https://doi.org/10.1016/S0001-6918(99)00050-5

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Ricciardi A, Cohen J (2007) The invasiveness of an introduced species does not predict its impact. Biological Invasions 9: 309–315. https://doi.org/10.1007/s10530-006-9034-4

Roy HE, Peyton J, Aldridge DC, Bantock T, Blackburn TM, Britton R, Clark P, Cook E, Dehnen-Schmutz K, Dines T, Dobson M, Edwards F, Harrower C, Harvey MC, Minchin D, Noble DG, Parrott D, Pocock MJO, Preston CD, Roy S, Salisbury A, Schönrogge K, Sewell J, Shaw RH, Stebbing P, Stewart AJA, Walker KJ (2014) Horizon scanning for invasive alien species with the potential to threaten biodiversity in Great Britain. Global Change Biology 20: 3859–3871. https://doi.org/10.1111/gcb.12603

Roy HE, Rabitsch W, Scalera R, Stewart A, Gallardo B, Genovesi P, Essl F, Adriaens T, Bacher S, Booy O, Branquart E, Brunel S, Copp GH, Dean H, D'hondt B, Josefsson M, Kenis M, Kettunen M, Linnamagi M, Lucy F, Martinou A, Moore N, Nentwig W, Nieto A, Pergl J, Peyton J, Roques A, Schindler S, Schönrogge K, Solarz W, Stebbing PD, Teodora T, Vanderhoeven S, van Valkenburg J, Zenetos A (2018) Developing a framework of minimum standards for the risk assessment of alien species. Journal of Applied Ecology 55: 526–538. https://doi.org/10.1111/1365-2664.13025

Sandvik H, Hilmo O, Finstad AG, Hegre H, Moen TL, Rafoss T, Skarpaas O, Elven R, Sandmark H, Gederaas L (2019) Generic ecological impact assessment of alien species

(GEIAA): the third generation of assessments in Norway. Biological Invasions 21: 2803–2810. https://doi.org/10.1007/s10530-019-02033-6

Schlaepfer MA, Sax DF, Olden JD (2011) The potential conservation value of non-native species. Conservation Biology 25: 428–437. https://doi.org/10.1111/j.1523-1739.2010.01646.x

Thompson PM, Hastie GD, Nedwell J, Barham R, Brookes KL, Cordes LS, Bailey H, McLean N (2013) Framework for assessing impacts of pile-driving noise from offshore wind farm construction on a harbour seal population. Environmental Impact Assessment Review 43: 73–85. https://doi.org/10.1016/j.eiar.2013.06.005

Turbé A, Strubbe D, Mori E, Carrete M, Chiron F, Clergeau P, González-Moreno P, Le Louarn M, Luna A, Menchetti M, Nentwig W, Parau LG, Postigo JL, Rabitsch W, Senar JC, Tollington S, Vanderhoeven S, Weiserbs A, Shwartz A (2017) Assessing the assessments: evaluation of four impact assessment protocols for invasive alien species. Diversity and Distributions 23: 297–307. https://doi.org/10.1111/ddi.12528

Vanderhoeven S, Branquart E, Casaer J, D'hondt B, Hulme PE, Shwartz A, Strubbe D, Turbé A, Verreycken H, Adriaens T (2017) Beyond protocols: improving the reliability of expert-based risk analysis underpinning invasive species policies. Biological Invasions 19: 2507–2517. https://doi.org/10.1007/s10530-017-1434-0

Van Der Wal R, Fischer A, Selge S, Larson BM (2015) Neither the public nor experts judge species primarily on their origins. Environmental Conservation 42: 349–355. https://doi.org/10.1017/S0376892915000053

Supplementary material I

Supplementary tables and figures

Authors: Ashley N. Schulz, Angela M. Mech, Craig R. Allen, Matthew P. Ayres, Kamal J.K. Gandhi, Jessica Gurevitch, Nathan P. Havill, Daniel A. Herms, Ruth A. Hufbauer, Andrew M. Liebhold, Kenneth F. Raffa, Michael J. Raupp, Kathryn A. Thomas, Patrick C. Tobin, Travis D. Marsico

Data type: statistical data

Explanation note: **Table S1.** List of the 15 participating assessors with their affiliation at the time of the 2016 HIWG meeting and their area(s) of expertise. **Table S2.** List of 58 conifer-specialist insect species and their nine-point (1–9 impact scale), ternary (1 = low, 2 = medium, 3 = high), and binary (0 = low, 1 = high) impacts as assessed by the initial assessor. Conifer-specialist insect species highlighted in gray were not included in the impact assessment, since they had no documented damage on their host plants in North America. **Table S3.** Directions on how to complete the impact assessment for non-native, conifer-specialist insects in North America, including scales for assessment of expertise and uncertainty. Developed by A.N. Schulz. **Table S4.** Level of expertise, impact score, and level of uncertainty assigned by the four assessors for each of the 41 conifer-specialist insects included in this study. **Figure S1.** Expected 95% confidence intervals for species-specific estimates of impact for 3–11 assessors per conifer-specialist insect species. **Figure S2.** Box plot depicting the distribution of assessor expertise among conifer-specialist insect species.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.55.38981.suppl1